```
   1 TCGCGGCGGC CGTGATGGCT GGTGACGGCG GGGCCGGGCA GGGGACCGGG
  51 GCCGCGGCCC GGGAGCGGGC CAGCTGCCGG GAGCCCTGAA TCACCGCCTG
 101 GCCCGACTCC ACCATGAACG TCGCGCTGCA GGAGCTGGGA GCTGGCAGCA
 151 ACATGGTGGA GTACAAACGG GCCACGCTTC GGGATGAAGA CGCACCCGAG
 201 ACCCCCGTAG AGGGCGGGGC CTCCCCGGAC GCCATGGAGG TGGGCAAGGG
 251 GGCTTCCCCT TTCTCACCAG GCCCCAGCCC TGGCATGACG CCTGGCACAC
 301 CCAGGAGCTC TGGGCTGTTC TGGAGGGTCA CCTGCCCCCA CCTCCGCTCC
 351 ATCTCTGGCC TCTGCTCTAG GACTATGGTG GGATTCCAGA AGGGGACAAG
 401 ACAGCTGTTA GGCTCACGCA CGCAGCTGGA GCTGGTCTTA GCAGGTGCCT
 451 CTCTACTGCT GGCTGCACTG CTTCTGGGCT GCCTTGTGGC CCTAGGGGTC
 501 CAGTACCACA GAGACCCATC CCACAGCACC TGCCTTACAG AGGCCTGCAT
 551 TCGAGTGGCT GGAAAAATCC TGGAGTCCCT GGACCGAGGG GTGAGCCCCT
 601 GTGAGGACTT TTACCAGTTC TCCTGTGGGG GCTGGATTCG GAGGAACCCC
 651 CTGCCCGATG GGCGTTCTCG CTGGAACACC TTCAACAGCC TCTGGGACCA
 701 AAACCAGGCC ATACTGAAGC ACCTGCTTGA AAACACCACC TTCAACTCCA
 751 GCAGTGAAGC TGAGCAGAAG ACACAGCGCT TCTACCTATC TTGCCTACAG
 801 GTGGAGCGCA TTGAGGAGCT GGGAGCCCAG CCACTGAGAG ACCTCATTGA
 851 GAAGATTGGT GGTTGGAACA TTACGGGGCC CTGGGACCAG GACAACTTTA
 901 TGGAGGTGTT GAAGGCAGTA GCAGGGACCT ACAGGGCCAC CCCATTCTTC
 951 ACCGTCTACA TCAGTGCCGA CTCTAAGAGT TCCAACAGCA ATGTTATCCA
1001 GGTGGACCAG TCTGGGCTCT TTCTGCCCTC TCGGGATTAC TACTTAAACA
1051 GAACTGCCAA TGAGAAAGTG CTCACTGCCT ATCTGGATTA CATGGAGGAA
1101 CTGGGGATGC TGCTGGGTGG GCGGCCCACC TCCACGAGGG AGCAGATGCA
1151 GCAGGTGCTG GAGTTGGAGA TACAGCTGGC CAACATCACA GTGCCCCAGG
1201 ACCAGCGGCG CGACGAGGAG AAGATCTACC ACAAGATGAG CATTTCGGAG
1251 CTGCAGGCTC TGGCGCCCTC CATGGACTGG CTTGAGTTCC TGTCTTTCTT
1301 GCTGTCACCA TTGGAGTTGA GTGACTCTGA GCCTGTGGTG GTGTATGGGA
1351 TGGATTATTT GCAGCAGGTG TCAGAGCTCA TCAACCGCAC GGAACCAAGC
1401 ATCCTGAACA ATTACCTGAT CTGGAACCTG GTGCAAAAGA CAACCTCAAG
1451 CCTGGACCGA CGCTTTGAGT CTGCACAAGA GAAGCTGCTG GAGACCCTCT
1501 ATGGCACTAA GAAGTCCTGT GTGCCGAGGT GGCAGACCTG CATCTCCAAC
1551 ACGGATGACG CCCTTGGCTT TGCTTTGGGG TCCCTCTTCG TGAAGGCCAC
1601 GTTTGACCGG CAAAGCAAAG AAATTGCAGA GGGGATGATC AGCGAAATCC
1651 GGACCGCATT TGAGGAGGCC CTGGGACAGC TGGTTTGGAT GGATGAGAAG
1701 ACCCGCCAGG CAGCCAAGGA GAAAGCAGAT GCCATCTATG ATATGATTGG
1751 TTTCCCAGAC TTTATCCTGG AGCCCAAAGA GCTGGATGAT GTTTATGACG
1801 GGTACGAAAT TTCTGAAGAT TCTTTCTTCC AAAACATGTT GAATTTGTAC
1851 AACTTCTCTG CCAAGGTTAT GGCTGACCAG CTCCGCAAGC CTCCCAGCCG
1901 AGACCAGTGG AGCATGACCC CCCAGACAGT GAATGCCTAC TACCTTCCAA
1951 CTAAGAATGA GATCGTCTTC CCCGCTGGCA TCCTGCAGGC CCCCTTCTAT
2001 GCCCGCAACC ACCCCAAGGC CCTGAACTTC GGTGGCATCG GTGTGGTCAT
2051 GGGCCATGAG TTGACGCATG CCTTTGATGA CCAAGGGCGC GAGTATGACA
2101 AAGAAGGGAA CCTGCGGCCC TGGTGGCAGA ATGAGTCCCT GGCAGCCTTC
2151 CGGAACCACA CGGCCTGCAT GGAGGAACAG TACAATCAAT ACCAGGTCAA
2201 TGGGGAGAGG CTCAACGGCC GCCAGACGCT GGGGGAGAAC ATTGCTGACA
2251 ACGGGGGGCT GAAGGCTGCC TACAATGCTT ACAAAGCATG GCTGAGAAAG
2301 CATGGGGAGG AGCAGCAACT GCCAGCCGTG GGGCTCACCA ACCACCAGCT
2351 CTTCTTCGTG GGATTTGCCC AGGTGTGGTG CTCGGTCCGC ACACCAGAGA
2401 GCTCTCACGA GGGGCTGGTG ACCGACCCCC ACAGCCCTGC CCGCTTCCGC
2451 GTGCTGGGCA CTCTCTCCAA CTCCCGTGAC TTCCTGCGGC ACTTCGGCTG
2501 CCCTGTCGGC TCCCCCATGA ACCCAGGGCA GCTGTGTGAG GTGTGGTAGA
2551 CCTGGATCAG GGGAGAAATG CCCAGCTGTC ACCAGACCTG GGGCAGCTCT
2601 CCTGACAAAG CTGTTTGCTC TTGGGTTGGG AGGAAGCAAA TGCAAGCTGG
2651 GCTGGGTCTA GTCCCTCCCC CCCACAGGTG ACATGAGTAC AGACCCTCCT
2701 CAATCACCAC ATTGTGCCTC TGCTTTGGGG GTGCCCCTGC CTCCAGCAGA
2751 GCCCCCACCA TTCACTGTGA CATCTTTCCG TGTCACCCTG CCTGGAAGAG
2801 GTCTGGGTGG GGAGGCCAGT TCCCATAGGA AGGAGTCTGC CTCTTCTGTC
2851 CCCAGGCTCA CTCAGCCTGG CGGCCATGGG GCCTGCCGTG CCTGCCCCAC
2901 TGTGACCCAC AGGCCTGGGT GGTGTACCTC CTGGACTTCT CCCCAGGCTC
2951 ACTCAGTGCG CACTTAGGGG TGGACTCAGC TCTGTCTGGC TCACCCTCAC
3001 GGGCTACCCC CACCTCACCC TGTGCTCCTT GTGCCACTGC TCCCAGTGCT
```

FIGURE 1A

```
 3051 GCTGCTGACC TTCACTGACA GCTCCTAGTG GAAGCCCAAG GGCCTCTGAA
 3101 AGCCTCCTGC TGCCCACTGT TTCCCTGGGC TGAGAGGGGA AGTGCATATG
 3151 TGTAGCGGGT ACTGGTTCCT GTGTCTTAGG GCACAAGCCT TAGCAAATGA
 3201 TTGATTCTCC CTGGACAAAG CAGGAAAGCA GATAGAGCAG GGAAAAGGAA
 3251 GAACAGAGTT TATTTTTACA GAAAAGAGGG TGGGAGGGTG TGGTCTTGGC
 3301 CCTTATAGGA CCCTGTGCCA ATAAACAGAC ATGCATCCGT CAAAAAAAAA
 3351 AAAAAAAAAA AAAAAAAAAA AAAAAAA  (SEQ ID NO:1)
```

FEATURES:

```
5'UTR:       1-113
Start Codon: 114
Stop Codon:  2547
3'UTR:       2550
```

Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA|18000005141003 /altid=gi|7662200 /def=ref|NP_055508.1| KIAA...	1550	0.0
CRA|18000005084162 /altid=gi|2136744 /def=pir||I46078 endotheli...	1430	0.0
CRA|18000005012182 /altid=gi|1706565 /def=sp|Q10711|ECE2_BOVIN ...	1430	0.0
CRA|150000075554683 /altid=gi|9789315 /def=gb|AAF98287.1|AF2302...	1027	0.0
CRA|1000682324124 /altid=gi|5821116 /def=dbj|BAA83687.1| (AB031...	1001	0.0
CRA|108000024636251 /altid=gi|12721007 /def=ref|XP_001827.2| en...	1001	0.0
CRA|18000004932659 /altid=gi|627989 /def=pir||A53679 endothelin...	1000	0.0
CRA|18000005060029 /altid=gi|1706564 /def=sp|P42893|ECE1_RAT EN...	996	0.0
CRA|18000005155376 /altid=gi|3287157 /def=emb|CAA19767.1| (AL03...	995	0.0
CRA|18000004985166 /altid=gi|1082351 /def=pir||JC2521 endotheli...	995	0.0

BLAST dbEST hits:

	Score	E
gi|6837875 /dataset=dbest /taxon=9606 /org=...	1094	0.0
gi|5925169 /dataset=dbest /taxon=9606 ...	983	0.0
gi|10725997 /dataset=dbest /taxon=96...	519	e-144
gi|2162041 /dataset=dbest /taxon=9606 ...	400	e-109

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

From BLAST dbEST hits:

```
gi|6837875   lung
gi|5925169   amygdala
gi|10725997  adrenal gland
gi|2162041   total fetus
```

From tissue screening panels:

Hippocampus

FIGURE 1B

```
   1 MNVALQELGA GSNMVEYKRA TLRDEDAPET PVEGGASPDA MEVGKGASPF
  51 SPGPSPGMTP GTPRSSGLFW RVTCPHLRSI SGLCSRTMVG FQKGTRQLLG
 101 SRTQLELVLA GASLLLAALL LGCLVALGVQ YHRDPSHSTC LTEACIRVAG
 151 KILESLDRGV SPCEDFYQFS CGGWIRRNPL PDGRSRWNTF NSLWDQNQAI
 201 LKHLLENTTF NSSSEAEQKT QRFYLSCLQV ERIEELGAQP LRDLIEKIGG
 251 WNITGPWDQD NFMEVLKAVA GTYRATPFFT VYISADSKSS NSNVIQVDQS
 301 GLFLPSRDYY LNRTANEKVL TAYLDYMEEL GMLLGGRPTS TREQMQQVLE
 351 LEIQLANITV PQDQRRDEEK IYHKMSISEL QALAPSMDWL EFLSFLLSPL
 401 ELSDSEPVVV YGMDYLQQVS ELINRTEPSI LNNYLIWNLV QKTTSSLDRR
 451 FESAQEKLLE TLYGTKKSCV PRWQTCISNT DDALGFALGS LFVKATFDRQ
 501 SKEIAEGMIS EIRTAFEEAL GQLVWMDEKT RQAAKEKADA IYDMIGFPDF
 551 ILEPKELDDV YDGYEISEDS FFQNMLNLYN FSAKVMADQL RKPPSRDQWS
 601 MTPQTVNAYY LPTKNEIVFP AGILQAPFYA RNHPKALNFG GIGVVMGHEL
 651 THAFDDQGRE YDKEGNLRPW WQNESLAAFR NHTACMEEQY NQYQVNGERL
 701 NGRQTLGENI ADNGGLKAAY NAYKAWLRKH GEEQQLPAVG LTNHQLFFVG
 751 FAQVWCSVRT PESSHEGLVT DPHSPARFRV LGTLSNSRDF LRHFGCPVGS
 801 PMNPGQLCEV W  (SEQ ID NO:2)
```

FEATURES:
Functional domains and key regions:

```
[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

Number of matches: 9
      1    207-210 NTTF
      2    211-214 NSSS
      3    252-255 NITG
      4    312-315 NRTA
      5    357-360 NITV
      6    424-427 NRTE
      7    580-583 NFSA
      8    673-676 NESL
      9    681-684 NHTA

[2] PDOC00004 PS00004 CAMP_PHOSPHO_SITE
cAMP- and cGMP-dependent protein kinase phosphorylation site

             18-21 KRAT

[3] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 8
      1      21-23 TLR
      2      62-64 TPR
      3    220-222 TQR
      4    272-274 TYR
      5    340-342 STR
      6    465-467 TKK
      7    582-584 SAK
      8    757-759 SVR

[4] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 19
      1      21-24 TLRD
      2      30-33 TPVE
      3    103-106 TQLE
      4    161-164 SPCE
```

FIGURE 2A

5	192-195	SLWD
6	212-215	SSSE
7	214-217	SEAE
8	314-317	TANE
9	340-343	STRE
10	376-379	SISE
11	398-401	SPLE
12	403-406	SDSE
13	445-448	SSLD
14	453-456	SAQE
15	478-481	SNTD
16	514-517	TAFE
17	613-616	TKNE
18	705-708	TLGE
19	763-766	SSHE

[5] PDOC00007 PS00007 TYR_PHOSPHO_SITE
Tyrosine kinase phosphorylation site

Number of matches: 4

1	365-372	RRDEEKIY
2	457-463	KLLETLY
3	535-542	KEKADAIY
4	555-561	KELDDVY

[6] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 14

1	9-14	GAGSNM
2	57-62	GMTPGT
3	61-66	GTPRSS
4	122-127	GCLVAL
5	159-164	GVSPCE
6	271-276	GTYRAT
7	331-336	GMLLGG
8	335-340	GGRPTS
9	464-469	GTKKSC
10	643-648	GVVMGH
11	714-719	GGLKAA
12	715-720	GLKAAY
13	782-787	GTLSNS
14	795-800	GCPVGS

[7] PDOC00047 PS00048 PROTAMINE_P1
Protamine P1 signature

776-787 ARFRVLGTLSNS

[8] PDOC00129 PS00142 ZINC_PROTEASE
Neutral zinc metallopeptidases, zinc-binding region signature

645-654 VMGHELTHAF

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	43	63	0.638	Putative
2	109	129	2.142	Certain
3	380	400	0.619	Putative
4	736	756	0.890	Putative

FIGURE 2B

BLAST Alignment to Top Hit:

```
>CRA|18000005141003 /altid=gi|7662200 /def=ref|NP_055508.1| KIAA0604
          gene product [Homo sapiens] /org=Homo sapiens
          /taxon=9606 /dataset=nraa /length=765
         Length = 765

 Score = 1550 bits (3969), Expect = 0.0
 Identities = 765/811 (94%), Positives = 765/811 (94%), Gaps = 46/811 (5%)

Query: 1   MNVALQELGAGSNMVEYKRATLRDEDAPETPVEGGASPDAMEVGKGASPFSPGPSPGMTP 60
           MNVALQELGAGSNMVEYKRATLRDEDAPETPVEGGASPDAMEV
Sbjct: 1   MNVALQELGAGSNMVEYKRATLRDEDAPETPVEGGASPDAMEV---------------- 43

Query: 61  GTPRSSGLFWRVTCPHLRSISGLCSRTMVGFQKGTRQLLGSRTQLELVLAGASLLLAALL 120
                                      GFQKGTRQLLGSRTQLELVLAGASLLLAALL
Sbjct: 44  ----------------------------GFQKGTRQLLGSRTQLELVLAGASLLLAALL 74

Query: 121 LGCLVALGVQYHRDPSHSTCLTEACIRVAGKILESLDRGVSPCEDFYQFSCGGWIRRNPL 180
           LGCLVALGVQYHRDPSHSTCLTEACIRVAGKILESLDRGVSPCEDFYQFSCGGWIRRNPL
Sbjct: 75  LGCLVALGVQYHRDPSHSTCLTEACIRVAGKILESLDRGVSPCEDFYQFSCGGWIRRNPL 134

Query: 181 PDGRSRWNTFNSLWDQNQAILKHLLENTTFNSSSEAEQKTQRFYLSCLQVERIEELGAQP 240
           PDGRSRWNTFNSLWDQNQAILKHLLENTTFNSSSEAEQKTQRFYLSCLQVERIEELGAQP
Sbjct: 135 PDGRSRWNTFNSLWDQNQAILKHLLENTTFNSSSEAEQKTQRFYLSCLQVERIEELGAQP 194

Query: 241 LRDLIEKIGGWNITGPWDQDNFMEVLKAVAGTYRATPFFTVYISADSKSSNSNVIQVDQS 300
           LRDLIEKIGGWNITGPWDQDNFMEVLKAVAGTYRATPFFTVYISADSKSSNSNVIQVDQS
Sbjct: 195 LRDLIEKIGGWNITGPWDQDNFMEVLKAVAGTYRATPFFTVYISADSKSSNSNVIQVDQS 254

Query: 301 GLFLPSRDYYLNRTANEKVLTAYLDYMEELGMLLGGRPTSTREQMQQVLELEIQLANITV 360
           GLFLPSRDYYLNRTANEKVLTAYLDYMEELGMLLGGRPTSTREQMQQVLELEIQLANITV
Sbjct: 255 GLFLPSRDYYLNRTANEKVLTAYLDYMEELGMLLGGRPTSTREQMQQVLELEIQLANITV 314

Query: 361 PQDQRRDEEKIYHKMSISELQALAPSMDWLEFLSFLLSPLELSDSEPVVVYGMDYLQQVS 420
           PQDQRRDEEKIYHKMSISELQALAPSMDWLEFLSFLLSPLELSDSEPVVVYGMDYLQQVS
Sbjct: 315 PQDQRRDEEKIYHKMSISELQALAPSMDWLEFLSFLLSPLELSDSEPVVVYGMDYLQQVS 374

Query: 421 ELINRTEPSILNNYLIWNLVQKTTSSLDRRFESAQEKLLETLYGTKKSCVPRWQTCISNT 480
           ELINRTEPSILNNYLIWNLVQKTTSSLDRRFESAQEKLLETLYGTKKSCVPRWQTCISNT
Sbjct: 375 ELINRTEPSILNNYLIWNLVQKTTSSLDRRFESAQEKLLETLYGTKKSCVPRWQTCISNT 434

Query: 481 DDALGFALGSLFVKATFDRQSKEIAEGMISEIRTAFEEALGQLVWMDEKTRQAAKEKADA 540
           DDALGFALGSLFVKATFDRQSKEIAEGMISEIRTAFEEALGQLVWMDEKTRQAAKEKADA
Sbjct: 435 DDALGFALGSLFVKATFDRQSKEIAEGMISEIRTAFEEALGQLVWMDEKTRQAAKEKADA 494

Query: 541 IYDMIGFPDFILEPKELDDVYDGYEISEDSFFQNMLNLYNFSAKVMADQLRKPPSRDQWS 600
           IYDMIGFPDFILEPKELDDVYDGYEISEDSFFQNMLNLYNFSAKVMADQLRKPPSRDQWS
Sbjct: 495 IYDMIGFPDFILEPKELDDVYDGYEISEDSFFQNMLNLYNFSAKVMADQLRKPPSRDQWS 554

Query: 601 MTPQTVNAYYLPTKNEIVFPAGILQAPFYARNHPKALNFGGIGVVMGHELTHAFDDQGRE 660
           MTPQTVNAYYLPTKNEIVFPAGILQAPFYARNHPKALNFGGIGVVMGHELTHAFDDQGRE
Sbjct: 555 MTPQTVNAYYLPTKNEIVFPAGILQAPFYARNHPKALNFGGIGVVMGHELTHAFDDQGRE 614

Query: 661 YDKEGNLRPWWQNESLAAFRNHTACMEEQYNQYQVNGERLNGRQTLGENIADNGGLKAAY 720
           YDKEGNLRPWWQNESLAAFRNHTACMEEQYNQYQVNGERLNGRQTLGENIADNGGLKAAY
Sbjct: 615 YDKEGNLRPWWQNESLAAFRNHTACMEEQYNQYQVNGERLNGRQTLGENIADNGGLKAAY 674
```

FIGURE 2C

```
Query: 721 NAYKAWLRKHGEEQQLPAVGLTNHQLFFVGFAQVWCSVRTPESSHEGLVTDPHSPARFRV 780
           NAYKAWLRKHGEEQQLPAVGLTNHQLFFVGFAQVWCSVRTPESSHEGLVTDPHSPARFRV
Sbjct: 675 NAYKAWLRKHGEEQQLPAVGLTNHQLFFVGFAQVWCSVRTPESSHEGLVTDPHSPARFRV 734


Query: 781 LGTLSNSRDFLRHFGCPVGSPMNPGQLCEVW 811
           LGTLSNSRDFLRHFGCPVGSPMNPGQLCEVW
Sbjct: 735 LGTLSNSRDFLRHFGCPVGSPMNPGQLCEVW 765  (SEQ ID NO:4)
```

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
CE00310	E00310 neutral_endopeptidase	456.3	2.6e-133	1
PF01431	Peptidase family M13	270.4	2.4e-77	1
CE00339	E00339 vitamin_D_receptor	4.3	1.1	1

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
CE00339	1/1	367	398	..	412	443	.]	4.3	1.1
PF01431	1/1	607	810	..	1	225	[]	270.4	2.4e-77
CE00310	1/1	140	811	.]	66	798	.]	456.3	2.6e-133

FIGURE 2D

```
   1 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
  51 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
1001 NNNNNNNNNN NNNNNNNNNN NNNNNNCACC TTAGACTTGA CAGGCCTGCT
1051 TAGTCGGACT CTAAAGCACC CCTTTGCTTT TCGTTAAATA TTGCTTGGTG
1101 TTAGTTTTTT TTCTCCTTGT AAATCTCCCA AATAAAACGG TTTGCTTTCC
1151 CCAAGTTAGA AGTGTTAGCA CGTCTTTTCT TTAAATATCT GTGCATGGCT
1201 GTTTTTTTCC CTGCCAATTT GTCACCATCT GTAACCCTCC CTTTATGAGA
1251 CGATCTGATG ACAGCAGTTA TCTTGGAGAG TAGAAGTGTG GTCTTGAAGC
1301 GCCATGGAAG AGTAGAGTCA GTGTATGCTG TGTGTGTGTG GAGTGTATGC
1351 TCCCCCTGCA CTTGGTGTGT GTACATACAG AAACACAGTG TGCGTGTGTG
1401 TTGGCTCTGG GTGTGTTGTG CGTGTGTACA CTGTGTGTGA GTATGCAGTG
1451 TGTGTACATT CTGTGGGCAT CTCGTGTGTG TGTGGACTGT GTGCTGGGCG
1501 TCGTGCCTGC CCGTGTCCTT GGCGCCTTGG CGTCTATGCG TTCTCTGCAC
1551 ATAGGTAGGT ACCACGTGCA CACCCTGAAT GTGAGTGAAC TGCCTGTGTG
1601 CTATGTATTT GCCGGCTGAA GAGGGGCTGT GTGGACTACT GGGGGAAGAC
1651 GTTCCTCANG AGGGCATAAT TTCTCTAAAG TGCTTAAAGG GGATGGAGAG
1701 AGCCTGAAAT TTGGGGGAAG TAGGCCAAGG AGTATTATCA ACGTCTGGGC
1751 CTGGTTGAAT TTCATTACTT TTCCTAGGAA AGTAAATTAT GGGTGGCTTG
1801 AAGGAGGGTG CTGCTGAGAT GGGGGGCGGA CCATGAAGCG TGGAGGGGTC
1851 TCCGGTGTTG CTGGAGGGCA GCTGGAGCCT GCGGAGAGCC TCGGCGCGCT
1901 CCTCCCTCTC CCCCACCCTC CCCCCACCCC GGGCGGGGCT CCGCGTGGGG
1951 CGGTGGACTC GGGCGGGGGG GGGGGCGGCC GCGGCCGAGC GGGGGTGCTG
2001 CGCGGCGGCC GTGATGGCTG GTGACGGCGG GGCCGGGCAG GGGACCGGGG
2051 CCGCGGCCCG GGAGCGGGCC AGCTGCCGGG AGCCCTGAAT CACCGCCTGG
2101 CCCGACTCCA CCATGAACGT CGCGCTGCAG GAGCTGGGAG CTGGCAGCAA
2151 CGTGAGTGGG GGCCCCGGGC TCCACGGGAG GGGACTGGGT GGAGGGGGAC
2201 GAGGCAGAGG GGTCGGCCGC GGAGGGGCAG GCGGTGCCCG GCTCGCGGAG
2251 GTAAGGCTGC CTCCCGGGCC TGGTGGAGGG GTGATAGAGA GACCCCGGGC
2301 CCGAGAGCAG GGCAGGTGGG AAGGGAAGGG CCCTCTTAGC AGGGCGGAGG
2351 GGTCCGCGAG GCAGGGAGCA CTGGGGCAGG GTCGTGGGCA AATAGCCCTC
2401 TCTGCCTGAC CTCGGTTGGC AACCCCGACT GTCTGGCAGA TGGTGGAGTA
2451 CAAACGGGCC ACGCTTCGGG ATGAAGACGC ACCCGAGACC CCCGTAGAGG
2501 GCGGGGCCTC CCCGGACGCC ATGGAGGTGG GCAAGGGGGC TTCCCCTTTC
2551 TCACCAGGCC CCAGCCCTGG CATGACGCCT GGCACACCCA GGAGCTCTGG
2601 GCTGTTCTGG AGGGTCATCT GCCCCCACCT CCGCTCCATC TCTGGCCTCT
2651 GCTCTAGGAC TATGGTGAGG CGATGCTAAG CCGTGACGTT GCACAAAACA
2701 GACTCAAGGC TCAACTCACT GGCTGGCCTC ATTGCCCCCG GGCCCAGAGT
2751 TAACCCTGTG GCTCTGAAAA CTGCCTGTGG CTTCACCCTC TGGTAATCTT
2801 GGATCCCTGC CCTGCATCTC AGTCACTCTC TGTCCCCCTG TGTTCCCCAG
2851 GTGGGATTCC AGAAGGGGAC AAGACAGCTG TTAGGCTCAC GCACGCAGCT
2901 GGAGCTGGTC TTAGCAGGTG CCTCTCTACT GCTGGCTGCA CTGCTTCTGG
2951 GCTGCCTTGT GGCCCTAGGG GTCCAGTACC ACAGAGGTAG GTGGGCCCAC
3001 ACTCTTCGTC AGTATTCATA ACTAGGGGTT CTGGAGGCCT AAGGGCCTCT
```

FIGURE 3A

3051 AAGATTTTCA CTTGTGGGAA CCAAGCCTTC CCTGCAGAAA AGCCCCCGGC
3101 TTTGCTTTCT CTTCCCAACC TTCCTGCTGT CATGGCCCTT GCAGAGTTTG
3151 CCTCTTCCAG ACAGACAGAC TGACAGTCTC CTACCCTCCG GCCATGTTCC
3201 CTACCACAGA CCCATCCCAC AGCACCTGCC TTACAGAGGC CTGCATTCGA
3251 GTGGCTGGAA AAATCCTGGA GTCCCTGGAC CGAGGGGTGA GCCCCTGTGA
3301 GGACTTTTAC CAGTTCTCCT GTGNNNNNNN NNNNNNNNNN NNNNNNNNNN
3351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3451 NNNNNNNNNC TTAGCAAATA GGCAGTGTCC CATGAATGAG GAAGTGGATG
3501 GTTCTGTGAA CACTCCCAGA GGGTGGGGAG GCAGAGAGCA GGGGACTATT
3551 GAGAAGTGCA GATGGGTTTG ATGGGGGCAG AACTCTGGGT ACAATGGAGG
3601 GCCGCTTCTC TGCACTCTGT TTGGAGCACT GTCGTGGTGT GGTAGACACC
3651 AGGGAGCCTG TACTGCTTAG ATATCCTTGG GTCTCCATGG ACAGGGAGAG
3701 GAAGCCACGG CTTGCTGTTT CAGACACTCT TCCTGGGTCT GCGTTAGCAG
3751 GACTGCTCAT TGACAAGGCA AGGAGAGAAA CCGAGCAAGG GCCAGGGACT
3801 CCCCCTCAGC AGTTAACGTA ATTGCCACCT GGATCCTGTG TTCTGCCCCA
3851 CAGAAAACAC CACCTTCAAC TCCAGCAGTG AAGCTGAGCA GAAGACACAG
3901 CGCTTCTACC TATCTTGCCT ACAGGTGGAG CGCATTGAGG AGCTGGGAGC
3951 CCAGCCACTG AGAGACCTCA TTGAGAAGGT AGGGCCACTG AGCCGGTTGA
4001 GGGCAGGGGA GCAGGAGAGG CCTTGAGAGA GGAGATGGCC CAGGAACGCT
4051 TTGGGAGCTC CTGCACTAAT CATTCCACTT ATGGTCTCTA CATAGATTGG
4101 TGGTTGGAAC ATTACGGGGC CCTGGGACCA GGACAACTTT ATGGAGGTGT
4151 TGAAGGCAGT AGCAGGGACC TACAGGGCCA CCCCATTCTT CACCGTCTAC
4201 ATCAGTGCCG ACTCTAAGAG TTCCAACAGC AATGTTATCC AGGTGATGAG
4251 CTGGGAAAGG GTGGGGAGAG ACTTAGGGAC ACTTTGCTGA GCCCAGACTT
4301 CCCTCTCCTG TGACAGGCAG GCTGGGCTGA CCCCCCGGCC CCACCCCTAC
4351 CCCCGCTCGG GAATTCAGGT TCCCATGGTG GGGAAAGCGA GGGGCTCACC
4401 TCCTTTCCTT GACATTGCAG GTGGACCAGT CTGGGCTCTT TCTGCCCTCT
4451 CGGGATTACT ACTTAAACAG AACTGCCAAT GAGAAAGTAA GGAACATCTT
4501 CCGAACCCCC ATCCCTACCC CTGGCTGAGC TGGGCTGATC CCTGTTGACT
4551 TTTCCCTTTG CCAAGGGTCA GAGCAGGGAA GGTGAGCCTA TCCTGTCACC
4601 TAGTGAACAA ACTGCCCCTC CTTTCTTTCT TCTTTTCTTC CTCCCTCCCT
4651 CCCTTTCTTC CCCTTTTCCT TCCTTCCTTC CTCTTATTCT TCTAGTAGGT
4701 TTCATAGACA CCTACTGTGT GCCAGGTCCA GTGGGGGAAT TCTGAGATAT
4751 AAGTTTNCCG AGCCCATTGC CAGCAGGAGA GGGGATCCTT TAGAGTCGCA
4801 CAAACAGGTC AGTCAAGTCT AAAGACNNNN NNNNNNNNNN NNNNNNNNNN
4851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
4951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5101 NNNNNNNNNN NNNNNNNNNN GCCTGNACTT GCATGCACCG CGGTTCGGCT
5151 NCTAGNAGNA TCCCCCCACT GCACTCCAGC CTGGGTGACG GAGAGAGACT
5201 CCGACTCAAA AAAAAAAAAA AAAAAGAAAG AAAAAGAAAG AAGGAACAGT
5251 TTAAACAAAA GTGTTGATGA GGCTGAGCAC AGTGGCTCAC ACCTGTAATC
5301 CCCGCACTTT GGGAGGCTGA GGCCGGCGGA TCACTTGAGG TTAGGAGTTC
5351 AAGACCAGGC TGGCCTACAA GGTGAAAACC CGTCTCTACT AAAAATACAA
5401 AAATTAGCCA GGCATGGTGG TGTGCACCTG TAATCTCAGC TACTTGGGAG
5451 GCTGAGGCAA AGAGAATCGC TTGAATCCAG GAGGCAGAGG TTGCAGTGAG
5501 CTGAGATGGC ACCACTGCAC TCCAGCCTGG GCAACAGAAC AAGACTTCAT
5551 CTCAAAAAAA AAAAAAAAAG TGTTGACGAG GGAAAGGCTA GGTGTGTCTG
5601 GACCATGGCA AGGGGTCCAC TGTGGTAAAA TATAGAACTC AAGGCAGATG
5651 AGAGGCTGGA GAGGTGGGCA GGAATGGGTT ATGGAGGGGA CCTTGAATAG
5701 CACACTACGG AGTTTATTCT GTAGCTCCCG GAGAGCCATT GCATGCTCCA
5751 AAGTAGGGAG GGAGCGCANT GCTTTGGGAA GTCAGTTTGT TTGGGGTGTG
5801 AAGAGTANAT GTGAGAACNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
5951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

FIGURE 3B

6101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
6951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
7951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
9101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

FIGURE 3C

```
 9151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 9951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
10951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
11401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNAG TTCCAGGCCC
11451 ACCCTTGGGC CAAACATGTT GAAGACCGCC ATGCTGTAGC TAGAACTTAC
11501 AAAAGATGTA AGCCTGGGCA TAGGTGGCCG GGTGCCGTTG TGGTCGCCAC
11551 GCTATCTTGG GGAGGGATTA AGGGCAAGGA AAATTCACCT TGAGGCCCAA
11601 GGAAGGCACA AGGGTTATCA CGTGAAGCCG AGGATCACCA TCACCATGCA
11651 CTAACACGCC TTGGGCAAGC ACGAAGCGAG GAGTTGCCAT CTCAAAACAA
11701 AAACGAAAAA CAAACAAACA AAATGCTAAT CAACTGTCAT TGGTAAGGCT
11751 TCTGGTCAAC AGTATGCTGT CAATAGTTAA GTTTTTGGGC TGGGCGCAGT
11801 GGCTCACGCC TGTAATCCCA GCACTTTGGG AGGCCAAAGC GGGTAGATCA
11851 CCTGAGGTCA GGAGTCGAGA CTAGCCTGGC CAACATGGCG AAACCCAGTC
11901 TCTACTAAAA ATATAAAAAT TAGCCAGGCG TGGTGGTGGG CACTTGTAAT
11951 CCCAGCTACT CAGGAAGCTG AGGCAGAACT GCTTGAACTG GGAAGTGGAG
12001 GTTGCAGTGA GCCGAGATCG TGCCATTGCA TTCCAGCCTG GGCGACAAGA
12051 GCAAAACTCC ATCTCAAAAA AAAAAAAAAA AAAAAAAGTT GTTTTTGGGG
12101 AGTCAAAAAT GAGGCCAGGC GCAGTGGCTC ATGCCTGTAA TCACAGCACT
12151 TTGGGAGGCC GAGGCGGGTG GATCACCTGA GGTCAGGAGT TCGTGACCAG
```

FIGURE 3D

12201 CTTGGCCAAC CTGGTGAAAC CCCGTCTCTA CTAAAAATAC AAAAATTAGC
12251 CGGGCATGGT GGCGGGCGCC CGTAATCTCA GCTACTTGGG CGGCTGAGGC
12301 AGGAGAATTG CTTCAACCCG GGAGGCAGAG GTTGCAATGA GCTGAGATCG
12351 CGCCACTGCA CTCCAGCCTT GGCGACAGAG GGAGACTCCA TGTCAAATTA
12401 AAAAAAAGAC CCCAGGATTT TGGACTGTGC AGGGGTCGGT GCCCCAAACC
12451 CCCACGTTGT TCAAGGTCAA CTGTACACTG TCATAGTCGG GAAAACTTCA
12501 TCACTGCAGC TGCTCCTGTT TCTTGAAACC TGAAGCGGGA AACTGGATCC
12551 TGGGACACTA CTGCCCCCTA TCGCCTGTTG GTCTTCAAAG AAATAATCCC
12601 TTCAATTTTG CAAGGCCTGT GGTGTCATTC CCTTTTAACA GATAAGGAAA
12651 CCGAGGCCAG GACGTGGTGG AAAATAATCA AGGTCACACA TCTATGTGCA
12701 AAAGTGGAGT AACAACCCAG GCTCCTCATT CCCAGGTCAG TCCAGTGACC
12751 TCAATTGACA TGAAATGTGT GAGGTCCTTC TGTGGCCCTG TGGCAGGGCC
12801 TGAAGAGGAC AGCGTATGTA AATCAAGTCT TGTGCCTTCA TGAGTGAGGC
12851 AGAGTAGAAA ATAACAGTAA TTCACTAGGA CCGAATCTGC ATTGTAAACA
12901 GAGAGGAAAG GGCTAGTATT TGGCAGAAGG ATGTCAAGGA ACATTTTAGA
12951 GATAAGAGGT GACATTTGGG TTCTGAGGGA TGAGTAGGAG TGTGCCAGGG
13001 TGCAAAGGAT GAAAAGACAG CTCTAGCAGC TGGTAAGGGC TAAGGGGCAT
13051 GGAGAAACAG CAAGACTTTG GGGAACTGGT AGAATTCTAA TTCTGGAAAA
13101 TTTGAACAAG GTAATTTTTT GTGTGTGGTT AAGGTATTAC ATACATACAG
13151 TAAAATAAAA TGCAATAGTT GCTGGGTGTG GAGGCTCACG CCTGTTAATC
13201 CCAGTACTTT GGAAGGCAGA GGCGGGTGGA TCATCTGAAG GTCAGGAGTT
13251 CGAGACCAGC CTGACCAACA TGGTGAAAAC CCGTCTCTAC TAAAAATACA
13301 AAAATTACCT GGGTGTGGTG GCAGGCGCCC GTAATCCCAG CTACTTGGGA
13351 GGCTAAGGGA GAAGAATAGC TTGAAACCCG GAGGTGGAGG TTGCAGTGAG
13401 CTGAGATTGC ACTATTGCGC TCCAGCCTGG GTGACAAGAG TGAAAAGCTG
13451 TCTCAAAATA AAATAAAAAT GTAATAGTCT AATTGATTTT TTTAAAAAAT
13501 GTAGACATCC ACGTATCTAC CACCTAGGTA AAGATACTAG AGATTCCAGC
13551 AACCTGGGAG GATCCCTCGT GCCCCTTTCA GGTCTATATG AGCCTCCACC
13601 GTTCCCCAGT CCCCTGGAAG GAGAGGGGGT GGGAGAGGCA ACATGAAACC
13651 TAAAAACCAG TGGGCTTCGC GCCTGTAATC CCAGCTATTG GGTTGGCTGA
13701 GGCAGGAGGA TCACTTGCCC AGGAGTTGGA GGCTGCAGTG AGCTATGATC
13751 GCGCCACCGC ACTCCAGCCT GGGCGACAGA TCAAGACCCC ATCTCTAAGC
13801 AAACAAACAA ATAAACACCC CTCAAAACCC ATGGCTTCAG GCCTGGCGCG
13851 GTAGCTTACT TCTGTAATCT CAGCACTTTG GGAGGCCGAG GAGGGCGGAT
13901 CACTTGAGGT CAGGAGTTCC AGACCAGACT GGCCAACATG GCGAAACCCC
13951 GTCTCTACTA AAAAATAAAA AAAAAAAAAA ATTGGCCGGG CGCGGTGGCT
14001 CACACCTGTA ATTACCAGCA GNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
14051 NNNNNNNNNN NNNNNNNNNN NTTTTAAAGA ATGTAGACAT CCACGTATCT
14101 ACCACCTAGG TAGAGATACT AGAGATTCCA GCAACCTGGG AGGATCCCTC
14151 GTGCGCCTTT CAGGTCTATA TGAGCCTCCA CCGTTCCCCA GTCCCCTGGA
14201 AGGAGAGGGG GTGGGAGAGG CAACATGAAA CCTAAAAACC AGTGGGCTTC
14251 GCGCCTGTAA TCCCAGCTAT TGGGTTGGCT GAGGCAGGAG GATCACTTGC
14301 CCAGGAGTTG GAGGCTGCAG TGAGCTATGA TCGCGCCACC GCACTCCAGC
14351 CTGGGCGACA GATCAAGACC CCATCTCTAA GCAAACAAAC AAATAAACAC
14401 CCCTCAAAAC CCATGGCTTC AGGCCTGGCG CGGTAGCTTA CTTCTGTAAT
14451 CTCAGCACTT TGGGAGGTCA AGGTGGGCGG ATCACTTGAA GTAAGGAGTT
14501 CAAGTACCAT CCTGGCTAAC ACGGTGAAAC CCCGTCTCTA CTGAAAAGAC
14551 AAAAAATTTA GCCGGGCGTG GTGGCGGGCG CCTTTAGTCT CAGCTACTCG
14601 GGAGGCTGAG GCAGGAGAAT GGCGTGAACC CGGGAGGTGG AGCTTGCAGT
14651 GAGCTGAGAT CGCACCACTG CACTCCAGTC TGGGTGACAG AGTGAGACTC
14701 CATCTCAAAA AAAAAAAAAA AGAAGTCAAA GTAGTAGAAA CTGCTGATAG
14751 ACTGAATGTG GGGGGTTAGG GAGATGGAGG AAGCTGAGTG ACTCCCAGGT
14801 TTCTTGCATG GGGGACTGAC TGGATATAAA ATTAGTTGTG GGCCGGGCAC
14851 GGTGGCTCAT GCCTTTAATC CCAGCACTTT GGGAGGCCAA AGCGGGCAGA
14901 TCACTTGAGC TCAGGAGTTC AAGACCAGCC TGGGAAACAT GGTGAGACCC
14951 CTTCTGTAAG GGNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
15001 NNNNNNNNNN NNNTGACCTT TTTTTGGCTC TGNTCGGTCA CTAGCANGCA
15051 AGTTATTGGG AGTCTACAAG ATTCTTTCAC ACTATGCCCT CAAAATTGAC
15101 TGTTCATGTA TGTGCAGACA TATAGAAAAA CAACGGGAGC CAGGCGCGGT
15151 GGCTCACGCC GGTAATCCCA GCACTTTGGG AGGCCAAGGC GGGTGAATCA
15201 TGGGGTCAGG AGTTCGAGAC CAGCCTGGCC AACATGGTGA AACCTGGTCT

FIGURE 3E

15251 CTACTAAAAA TACAAAAAAT TAGCCGGGCG TGGTGGCGGG TGTCTGTAAT
15301 CCCAGCTACT TGGGAGGCTG AGGCAGGAGA ATCACTTGAA CCCAGGAGGC
15351 GGAGGTTGCA GTGAGCCGAG ATCGCGCCAG TGCACTCCAG CCTGGGCGAC
15401 AGAGCAAAAC TCTGTCTCAA AAAAAAAAAA AAAAAAAAGA AAAGAAAAGA
15451 AAAACAACTG GATGTAAATT GATGAACAAA TGAAGTAGTG CTGCTTTGGG
15501 CAGTGGGATT ATAAGAGTCC TTTAAAGTTG TCTATGTGTT TATGTTTAAC
15551 TATATAACTA GAAGAAATAT TTATTTATTA GGATATGATA ATGGATGTGC
15601 TTAAAGTATT ACCTGTAAGG ATGTTTATGG TTTTTATGGC AATGTTGTTT
15651 ATAATAGCAG AAAATGAGAA CAGGTTAAAT GTCCAACTAT AGGGTAAAGG
15701 AAAAATAAAT TGTGGTTAGG ATGGGTTGTG AGGATCCTTA AATGGCTGAT
15751 ATATCTTTCA GCAAAAAAAG TAGGTTACAA AAAATATATA CCCTATACAA
15801 CATAATTCCA TATTTTATAT GCATATCAGG GGAGGGAAAA ACTCTAGAAG
15851 TGGGTAATCA AAATGTTAAA AGAACTTATC TATGAATGAG TGCTTTATAA
15901 CTGGTCTGTT CTTCAATTCT CAATTTTCCA AATTTTCTGT GAATGTCCTC
15951 TTTTCATAAT CAGATAAAAA TCATTGCACT AGGCTGGGCG TGGTGGTTCA
16001 CGCTTGTAAT CCCAGCACTT TGGGAGGCTG AGGCGGGTGG ATCACGTGGT
16051 CAGGAGTTCA AGACCAACCT GGCCAAGATG GTGAAACCCC AGCTCTACTA
16101 AAAATACAAA AATTACCCGG GCATGATGGC GGGAGCCTGT AATCCTAGCT
16151 ACTTGGGAGG CTGAGGCAGG AGAATCGCTT GAACTCGGGA GGCGGAGGTT
16201 GCAGTGAGCC GAGATTGCGC CACTGCACTC CATCCTAGGT AACACAGCCA
16251 GACTCTGTCT CAAAAAAAAA AAAAAATCAT TGCACTATAT TAAATTATAA
16301 TATAATTTGA TGAACTTATT GTCAATTAAA ATGTGTACTT AATTAAGAAA
16351 AAAGCCAGCC ACAATCCCAG TACCTTTACA AATGGTGTTT CCTTCTCATC
16401 GTCTCCAGGT GCTCAGCCGT ATTTCTTTAG TCTAGACGTT CCCATTTCCC
16451 CTGGGTGGAC AGGGATGGGG CACCAAGGGT GGATGGGTGG GGCAGGGATG
16501 CATTCAGTGC AGGGGAAGGC TGACTTTACC TCCTCCCTCC CAGGCAGAGG
16551 GGATGATCAG CGAAATCCGG ACCGCATTTG AGGAGGCCCT GGGACAGCTG
16601 GTTTGGATGG ATGAGAAGAC CCGCCAGGCA GCCAAGGAGA AAGTGAGCGG
16651 TGGCTAGGGT TGGGGCGCCA TCTTGAGGTG GGGTTCAAGG ATACAGTTTT
16701 GCTAGGAACC TGGGGAAGGA AACAAACCCT TAACCTGGTC TCTTCAGGCA
16751 GATGCCATCT ATGATATGAT TGGTTTCCCA GACTTTATCC TGGAGCCCAA
16801 AGAGCTGGAT GATGTTTATG ACGGGGTGAG TACCTACGCT CATCAGTACT
16851 GAACTTCAGC CCTGTAGAGG GCACTGTTCC CTGGGCTTAG AAATTGGGGC
16901 TCAAGCACTG GGAAAGAGGT GCTTGTCGGT TTCTTTTAGA GGCAGATGGA
16951 GGTAACCAGC ATTGTTAAAA TGTTGGCTCT GTGACAGGCT GCAGGCCAAA
17001 CAGCAGTGAA ATATAGTGCT AACGAGCCAA GATTTGGAGT CAAGCCTAAT
17051 CAAATTCTGT TTCTACCTCT AACTTTGTAA CCTTAACAAA ATCTCTCTAG
17101 GCCTTGGTTT CATTTTCTGT AAAATGGGGG TCCTACTAGT GCCTTCCTCA
17151 TAGGGTTGTT GTGAGATAAA TGAATACAGT ATGTAAAAAA ACAGCACCCA
17201 TAACATAAAT GGCCTTTAAA TATTGCCAAT TATGGTTTAC TAGATATTTT
17251 ACAGTTGAGG AAACTGAGGT TTGGAGAGAT ACTAATGAGT AGCCAAACTG
17301 GCGCTATTAT CTTCTCCAAT GGATTCTCTT GCTCTCTGTC TACTTCCCAA
17351 CTTACCACAG AACAAANNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

FIGURE 3F

```
18301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18801 NNNNNNNNNN NNNNNNNNNN NNNNNAGAAT CACCAACAGC ATTGGATGAA
18851 AATAAAGAAG AACAAGAGGT TCGTTTGAGA GGAAGCCGGG AAAATTCTCT
18901 CGATAAAGAA ATGCAAGTGC GCGCGCGGCG CAACCACTAC AATAGTGTGT
18951 CGTCCACCCC AGAGAGTGAA GGGGGCCCCC CCCGCCCCAA AGGAAAGGGG
19001 TAGTGTCCAC GCCGCTCCAC AAAGAGAGAG AAGGAAAGAA GTAGTTTTCC
19051 CCCCCCCGGG GAGAAACCTT GGATGGGGCT CANCCCCCCC TCTTTTTTTT
19101 TCCCGCGAAA ACCCCCCCCA AAAAGTTTTT TTTAAAAAAC AAAAAAGGGG
19151 GGTTTGGTTT TTTGGGCCCC GTGGCCCCTT TGGTTTAAAT TGGGAGAAAG
19201 AGGGCTTAAA GGGGGGATTC AAGAAAAAAC CCCCCCCCAA TTGCCCCAAA
19251 TTGTAATTTC CTAACCCCAA AAGGGGCCCC TAAAATTTCC GGGGAAACCC
19301 GTGTGGGCAA TGGCCCATTA GTTTACCCAA TGCCTTTATT GACAAAGGTA
19351 GGGCCCCATG GAGTCGTCCC CTCTAGCCTA GAATTCCCAG TGGCTCCTGC
19401 AAGGGCCTTG GGACATTGAT GTAGCCCCAA GGGCCCTGAA GTCTGTGGAC
19451 CAGGGCTGGT GGGGCACTGC TGCCCCCAAG AGACGAGCTC TGGTTTTGGT
19501 GGGGTGCAAA GGTGAGTTCT CCTCAGGGCG CGAGTATGAC AAAGAAGGGA
19551 ACTGCGGCCC TGGTGGCAGA ATGAGTCCCT GGCAGCCTTC CGGAACCACA
19601 CGGCCTGCAT GGAGGAACAG TACAATCAAT ACCAGGTCAA TGGGGAGAGG  (SEQ ID NO:3)
```

FEATURES:

Start:	2113
Exon:	2113-2151
Intron:	2152-2439
Exon:	2440-2664
Intron:	2665-2850
Exon:	2851-2986
Intron:	2987-3209
Exon:	3210-3320
Intron:	3321-3519
Exon:	3520-3636
Intron:	3637-3853
Exon:	3854-3978
Intron:	3979-4095
Exon:	4096-4242
Intron:	4243-4420
Exon:	4421-4486
Intron:	4487-4576
Exon:	4577-4581
Intron:	4582-4707
Exon:	4708-4811
Intron:	4812-5525
Exon:	5526-5591
Intron:	5592-11594
Exon:	11595-11742
Intron:	11743-13150
Exon:	13151-13240
Intron:	13241-16408
Exon:	16409-16642
Intron:	16643-16747
Exon:	16748-16825
Intron:	16826-18962

FIGURE 3G

Exon: 18963-19222
Intron: 19223-19321
Exon: 19322-19347
Intron: 19348-19526
Exon: 19527-19647

CHROMOSOME MAP POSITION:
Chromosome # 3

ALLELIC VARIANTS (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
2707	A	G	Intron			
4209	C	T	Exon	289	S	S
4355	-	T G	Intron			
15455	-	A C	Intron			

Context:

DNA Position

2707
TGACCTCGGTTGGCAACCCCGACTGTCTGGCAGATGGTGGAGTACAAACGGGCCACGCTT
CGGGATGAAGACGCACCCGAGACCCCCGTAGAGGGCGGGGCCTCCCCGGACGCCATGGAG
GTGGGCAAGGGGGCTTCCCCTTTCTCACCAGGCCCCAGCCCTGGCATGACGCCTGGCACA
CCCAGGAGCTCTGGGCTGTTCTGGAGGGTCATCTGCCCCCACCTCCGCTCCATCTCTGGC
CTCTGCTCTAGGACTATGGTGAGGCGATGCTAAGCCGTGACGTTGCACAAAACAGACTCA
[A,G]
GGCTCAACTCACTGGCTGGCCTCATTGCCCCCGGGCCCAGAGTTAACCCTGTGGCTCTGA
AAACTGCCTGTGGCTTCACCCTCTGGTAATCTTGGATCCCTGCCCTGCATCTCAGTCACT
CTCTGTCCCCCTGTGTTCCCCAGGTGGGATTCCAGAAGGGGACAAGACAGCTGTTAGGCT
CACGCACGCAGCTGGAGCTGGTCTTAGCAGGTGCCTCTCTACTGCTGGCTGCACTGCTTC
TGGGCTGCCTTGTGGCCCTAGGGGTCCAGTACCACAGAGGTAGGTGGGCCCACACTCTTC

4209
CCTATCTTGCCTACAGGTGGAGCGCATTGAGGAGCTGGGAGCCCAGCCACTGAGAGACCT
CATTGAGAAGGTAGGGCCACTGAGCCGGTTGAGGGCAGGGGAGCAGGAGAGGCCTTGAGA
GAGGAGATGGCCCAGGAACGCTTTGGGAGCTCCTGCACTAATCATTCCACTTATGGTCTC
TACATAGATTGGTGGTTGGAACATTACGGGGCCCTGGGACCAGGACAACTTTATGGAGGT
GTTGAAGGCAGTAGCAGGGACCTACAGGGCCACCCCATTCTTCACCGTCTACATCAGTGC
[C,T]
GACTCTAAGAGTTCCAACAGCAATGTTATCCAGGTGATGAGCTGGGAAAGGGTGGGGAGA
GACTTAGGGACACTTTGCTGAGCCCAGACTTCCCTCTCCTGTGACAGGCAGGCTGGGCTG
ACCCCCCGGCCCCACCCCTACCCCCGCTCGGGAATTCAGGTTCCCATGGTGGGGAAAGCG
AGGGGCTCACCTCCTTTCCTTGACATTGCAGGTGGACCAGTCTGGGCTCTTTCTGCCCTC
TCGGGATTACTACTTAAACAGAACTGCCAATGAGAAAGTAAGGAACATCTTCCGAACCCC

4355
GAGCTCCTGCACTAATCATTCCACTTATGGTCTCTACATAGATTGGTGGTTGGAACATTA
CGGGGCCCTGGGACCAGGACAACTTTATGGAGGTGTTGAAGGCAGTAGCAGGGACCTACA
GGGCCACCCCATTCTTCACCGTCTACATCAGTGCCGACTCTAAGAGTTCCAACAGCAATG
TTATCCAGGTGATGAGCTGGGAAAGGGTGGGGAGAGACTTAGGGACACTTTGCTGAGCCC
AGACTTCCCTCTCCTGTGACAGGCAGGCTGGGCTGACCCCCCGGCCCCACCCCTACCCCC
[-,T,G]
CTCGGGAATTCAGGTTCCCATGGTGGGGAAAGCGAGGGGCTCACCTCCTTTCCTTGACAT
TGCAGGTGGACCAGTCTGGGCTCTTTCTGCCCTCTCGGGATTACTACTTAAACAGAACTG
CCAATGAGAAAGTAAGGAACATCTTCCGAACCCCCATCCCTACCCCTGGCTGAGCTGGGC
TGATCCCTGTTGACTTTTCCCTTTGCCAAGGGTCAGAGCAGGGAAGGTGAGCCTATCCTG
TCACCTAGTGAACAAACTGCCCCTCCTTTCTTTCTTCTTTTCTTCCTCCCTCCCTCCCTT

FIGURE 3H

```
15455   CACGCCGGTAATCCCAGCACTTTGGGAGGCCAAGGCGGGTGAATCATGGGGTCAGGAGTT
        CGAGACCAGCCTGGCCAACATGGTGAAACCTGGTCTCTACTAAAAATACAAAAAATTAGC
        CGGGCGTGGTGGCGGGTGTCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCA
        CTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCAGTGCACTCCAGCCTG
        GGCGACAGAGCAAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAGAAAAGAAAAGAAAAA
        [-,A,C]
        AACTGGATGTAAATTGATGAACAAATGAAGTAGTGCTGCTTTGGGCAGTGGGATTATAAG
        AGTCCTTTAAAGTTGTCTATGTGTTTATGTTTAACTATATAACTAGAAGAAATATTTATT
        TATTAGGATATGATAATGGATGTGCTTAAAGTATTACCTGTAAGGATGTTTATGGTTTTT
        ATGGCAATGTTGTTTATAATAGCAGAAAATGAGAACAGGTTAAATGTCCAACTATAGGGT
        AAAGGAAAAATAAATTGTGGTTAGGATGGGTTGTGAGGATCCTTAAATGGCTGATATATC
```

FIGURE 3I